Again, the above are merely illustrative of the many possible applications of the present invention, and it is expected that many more web-based enterprises, as well as other consumer applications (such as intelligent, interactive toys) can utilize the present teachings. Although the present invention has been described in terms of a preferred embodiment, it will be apparent to those skilled in the art that many alterations and modifications may be made to such embodiments without departing from the teachings of the present invention. It will also be apparent to those skilled in the art that many aspects of the present discussion have been simplified to give appropriate weight and focus to the more germane aspects of the present invention. The microcode and software routines executed to effectuate the inventive methods may be embodied in various forms, including in a permanent magnetic media, a non-volatile ROM, a CD-ROM, or any other suitable machine-readable format. Accordingly, it is intended that the all such alterations and modifications be included within the scope and spirit of the invention as defined by the following claims.

What is claimed is:

1. An interactive system adapted for responding to speech-based queries concerning topics addressed by such interactive system, the system comprising:

a query file for storing a plurality of topic query entries, each topic query entry including a query relating to one or more of the topics covered by the speech-based interactive system; and

an answer file for storing a plurality of topic answer entries, each topic answer entry including an answer to one or more of said plurality of topic query entries, such that each topic query entry has at least one associated topic answer entry; and

a speech recognition system for generating recognized speech data from partially processed speech data associated with a speech-based query concerning one of said topics, said partially processed speech data being received from a remote speech capturing system; and

said speech recognition system further cooperating with a natural language engine which generates recognized speech sentence data corresponding to said speech-based query using both semantic decoding and statistical based processing performed on said recognized speech utterance data;

a query formulation system for converting said recognized speech data into a search query suitable for identifying a topic query entry corresponding to said speech-based query, and for locating at least one topic answer entry best matching said speech-based query.

2. The system of claim 1, wherein said remote speech capturing system is located at a client site, and said speech recognition system is distributed across said client site and a separate server site.

3. The system of claim 1, wherein said speech recognition system is comprised of a first portion at a client based computing system for performing first signal processing operations on a speech input signal to create said partially processed speech data, and a second portion at a server based computing system for performing a second signal processing operation for completing processing of said partially processed speech data.

4. The system of claim 1, wherein said statistical processing is based on calculating noun phrases in said recognized speech utterance data.

5. The system of claim 1, wherein said statistical processing is further based on calculating verb phrases, adjective phrases and adverb phrases in said recognized speech utterance data.

6. The system of claim 1, wherein said semantic decoding is based on a term frequency calculation, which term frequency calculation is based on calculating a lexical distance between each word in said recognized speech utterance data with each word of one or more topic query entries.

7. The system of claim 6, wherein said one or more topic query entries are retrieved by an SQL search.

8. The system of claim 6, wherein semantic decoding further includes a coverage calculation, which coverage calculation is based on identifying a percentage of a number of terms in said recognized speech utterance data which appear in each of said one or more topic query entries.

9. The system of claim 6, wherein semantic decoding further includes a semantic similarity calculation, which semantic similarity calculation is based on identifying a semantic distance between two like parts of speech of said recognized speech utterance data and each of said one or more topic query entries.

10. The system of claim 1, wherein said query formulation system uses context parameters for recognizing said speech-based query.

11. A method of implementing a speech-based interactive query system, including the steps of:

(a) storing a plurality of topic query entries, each topic query entry including a query relating to one or more of topics covered by the speech-based interactive query system; and

(b) storing a plurality of topic answer entries, each topic answer entry including an answer to one or more of said plurality of topic query entries, such that each topic query entry has at least one associated topic answer entry; and

(c) generating recognized speech utterance data associated with a speech-based query concerning one of said topics, such that said recognized speech utterance data is generated by partial recognition processing of said speech-based query by a first signal processing routine executing at a first computing device, and then completing recognition of said speech-based query through processing performed by a second signal processing routine executing at a second computing device; and

(d) converting said recognized speech utterance data with a natural language process into recognized speech sentence data, said recognized speech data being used by a search query suitable for identifying a topic query entry corresponding to said speech-based query; and

(e) locating at least one topic answer entry best matching said speech-based query;

wherein said natural language process includes both semantic decoding and statistical based processing operations performed on said recognized speech utterance data.

12. The method of claim 11, wherein said first computing device is located at a client site, and said second computing device is located at a separate server site.

13. The method of claim 11, wherein said semantic decoding includes a term frequency calculation, which term frequency calculation is based on calculating a lexical distance between each word in said recognized speech utterance data with each word of one or more topic query entries.

14. The method of claim 13, wherein said semantic decoding further includes a coverage calculation, which coverage calculation is based on identifying a percentage of a number of terms in said recognized speech utterance data which appear in each of said one or more topic query entries.

15. The method of claim 14, wherein said semantic decoding further includes a semantic similarity calculation, which semantic similarity calculation is based on identifying a semantic distance between two like parts of speech of said recognized speech utterance data and each of said one or more topic query entries.

16. The method of claim 11, wherein during step (d) context parameters are used for formulating said search query, and said context parameters are used for dynamically determining and loading an appropriate grammar and dictionary file to be used for said speech-based query.

17. The method of claim 11, wherein said response undergoes a text to speech process so that said topic answer entry is expressed in audible form to a user.

18. A method of performing semantic decoding of a user question presented in a speech utterance to a networked online natural language processing system to determine a meaning of the user question, the method comprising the steps of:

(a) receiving a set of words forming the user question at the networked online natural language processing system;

wherein the user question can be associated with one or more known queries, and which known queries have a meaning that is understood by the networked online natural language processing system;

(b) determining a frequency of occurrence of one or more terms identified in the user question and in each of said one or more known queries;

(c) determining a percentage of terms which appear in the user question that also appear in each of said one or more known queries;

(d) measuring a semantic similarity between said set of words in the user question, and a corresponding set of words in each of said one or more known queries;

(e) combining the results of steps (b) through (d) to identify which one of said one or more known queries is to be used in identifying the meaning of the user question.

19. The method of claim 18, wherein each of steps (b) through (d) includes an associated weighting factor.

20. The method of claim 19, wherein a composite semantic metric is calculated based on steps (b) through (d) and a corresponding set of weighting factors.

21. The method of claim 18, wherein during step (c) a semantic similarity W between the user question (UQ) and each of said one or more known queries is calculated using WordNet, and by computing a semantic distance between UQ (q) and a known query (d) as follows:

$$\text{Sem}\,(T_{uq}, T_r) = [I(uq,r) + I(r,uq)] / [\text{Abs}[T_{uq}] + \text{Abs}[T_r]]$$

where I(uq,r) and I(r,uq) are values corresponding to the inverse semantic distances computed at a given sense and level of WordNet in both directions, and T is a term frequency metric.

22. The method of claim 18, wherein said semantic decoding is performed by the natural language processing system only when a statistical based processing operation first determines that two or more known queries can correspond to the user question.

23. The method of claim 22, wherein said statistical based processing operation is based on an analysis of noun phrases and other parts of speech.

24. The method of claim 18, wherein during step (d) a word by word comparison is made between each word in the user question and each word in each of said one or more known queries to compute a matrix of values identifying a minimum lexical distance between such words.

25. The method of claim 18, wherein the semantic decoding is performed by a natural language engine operating in part on an online server connected to the Internet and in part on a client device.

26. A method of populating a natural language speech lattice with semantically variant questions, the method comprising the steps of:

(a) receiving a user question;

(b) dividing the user question into a plurality of words corresponding to the user question;

(c) determining synonyms for each word in said plurality of words;

(d) formulating a random set of questions based on said synonyms;

(e) performing semantic decoding on said random set of questions, to identify a disambiguated set of questions;

(f) storing said set of disambiguated questions in a speech recognition lattice;

wherein said set of disambiguated questions correspond to semantic variants of questions that can be posed to a natural language speech engine.

27. The method of claim 25, wherein steps (a) through (f) are performed by a software program embodied in a machine readable media.

28. The method of claim 25, wherein the method is performed by a natural language engine operating on an online server connected to the Internet.